NPS-OR-01-007

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Longitudinal Studies Relating to Training Dead Time

Final Report

by

R.R. Read

June 2001

Approved for public release; distribution is unlimited.

Prepared for:
Deputy Chief of Naval Operations
N-8, Office of the Chief of Naval Operations
2000 Navy Pentagon
Washington, DC 20350-2000

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CA 93943-5000

RADM David R. Ellison
Superintendent

Richard Elster
Provost

This report was prepared for and funded by the Deputy Chief of Naval Operations.

This report was prepared by:

R.R. Read
Professor of Operations Research

Reviewed by:

R. KEVIN WOOD
Associate Chairman for Research
Department of Operations Research

JAMES N. EAGLE
Chairman
Department of Operations Research

Released by:

DAVID W. NETZER
Associate Provost and Dean of Research

REPORT DOCUMENTATION PAGE			Form approved OMB No 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.			
1. AGENCY USE ONLY (Leave blank)	**2. REPORT DATE** June 2001	**3. REPORT TYPE AND DATES COVERED** Final Report	
4. TITLE AND SUBTITLE Longitudinal Studies Relating to Training Dead Time			**5. FUNDING** N0003999WRDR037
6. AUTHOR(S) Robert R. Read			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Operations Research Department 1411 Cunningham Road Monterey, CA 93943-5219			**8. PERFORMING ORGANIZATION REPORT NUMBER** NPS-OR-01-007
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) N-8, Office of the Chief of Naval Operations 2000 Navy Pentagon Washington, DC 20350-2000			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES Final Report			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			**12b. DISTRIBUTION CODE**

13. ABSTRACT (Maximum 200 words.)

Recent studies have provided quantitative information relating to the very high cost of dead time (time that sailors are not undergoing training although assigned for training) in the Navy training system. These studies are based upon quarterly and monthly average on board (AOB) data, which provided the period averages for numerous categories of dead time and non-dead time. Data of this type are readily accessible. It has been suggested that a different data structure (i.e., longitudinal data, which records the time spent by sailors in the various categories measured from the beginning of the courses), would provide sharper information about what is happening and help to better understand the nature of the problems, their relative importance, and suggest types of remedial action. The present report presents some models of the longitudinal type and fits them to data. Specifically, it treats the holding time distributions measured from the beginning of a course until the entrance into a non-training state for academic attritions, academic setbacks, and interrupted instruction of the non-legal holiday type. Analysis shows that there is considerable variability of these distributions from course to course and year to year.

Also considered are the data needs for the longitudinal study of the downtimes between courses in a pipeline of courses.

14. SUBJECT TERMS Training, Setbacks, Attritions, Non-Homogeneous Poisson Processes			**15. NUMBER OF PAGES** 21
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	**18. SECURITY CLASSIFICATION OF THIS PAGE** UNCLASSIFIED	**19. SECURITY CLASSIFICATION OF ABSTRACT** UNCLASSIFIED	**20. LIMITATION OF ABSTRACT** Unlimited

NSN 7540-01-280-5800

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std 239-18

Abstract

Recent studies have provided quantitative information relating to the very high cost of dead time (time that sailors are not undergoing training although assigned for training) in the Navy training system. These studies are based upon quarterly and monthly average on board (AOB) data, which provided the period averages for numerous categories of dead time and non-dead time. Data of this type are readily accessible. It has been suggested that a different data structure (i.e., longitudinal data, which records the time spent by sailors in the various categories measured from the beginning of the courses), would provide sharper information about what is happening and help to better understand the nature of the problems, their relative importance, and suggest types of remedial action.

The present report discusses the data needs for this type of study and draws attention to the data acquisition problems. A limited amount of longitudinal data was acquired for selected courses and years. Models were constructed describing the number of days from entering a course until either academic attrition, academic setback, or interrupted instruction of the non-legal holiday type. A distance measure was developed for deciding the separation of one model fit from another. Its use shows that there is considerable variability of these distributions from course to course and from year to year.

Also considered are the data needs for the longitudinal study of the downtimes between courses in a pipeline of courses.

Keywords: Training, Setbacks, Attritions, Non homogeneous Poisson Processes

1. Introduction

The costs associated with training dead time have attracted an increasing amount of attention, [Rhoades, 1999]. Dead time in naval training schools refers to man-days lost when sailors assigned to schools are not in an instructional mode. There are many reasons for this condition. The broad categories of dead time are awaiting instruction (AI), interrupted instruction (II), and awaiting transfer (AT). The first, AI, is caused largely by sailors arriving for instruction prior to the convening of the course and/or the condition that space in the classroom is not available. The second, II, reflects a large number of seemingly random events that take students out of the classroom. This includes the legal holidays and these appear to be the more prominent contributors, although they are scheduled rather than random. The third, AT, often reflects glitches in the cutting of orders and the budgeting of PCS (permanent change of station) funds. These major forms have received much attention (see references). A nice description of the flow can be found in [Belcher, et al., 1999]

The data structures used in the cited studies of dead time are gathered at fixed points in time. Quarterly data are readily accessible from the Navy Integrated Training Resources and Administration System (NITRAS), but monthly data can be obtained upon request. The values are "average on board" (AOB) for the period; that is, the time average of the personnel count in a particular category for the period of time. The inferences are based upon these. Personnel in Manpower, Personnel, Training and Education (N813) have suggested that there may be important complementary information in the "holding time" distributions that measure the number of days that students stay in a specific state (category) prior to changing to another state. Such distributions are commonly called longitudinal.

There are two main kinds of states: Under Instruction (UI) and Not Under Instruction (NUI). The former is the preferred state for all sailors associated with a training status. The latter is the all-inclusive dead time state and contains, of course, AI, II, and AT as sub-states. It is important to reduce the holding times in these latter states. The author has been asked to look into the distributions of holding times. The goal is to identify explanatory variables, be they courses, seasons or policies that promote uptime (UI) and diminish down time (NUI).

The progress has been modest. The acquisition of appropriate data is difficult; the databases are not organized for direct access to such distributions. Some models for certain kinds of uptime have been developed and tested. The successful ones are coarse in nature. The proper data requirements are not yet fully developed. The present report documents the issues and clarifies the processes involved. Some modeling for the random times (i.e., due to attrition, setbacks and non- holiday II), until entry into an NUI state from an UI state has taken place. These are presented and tested in the report. The levels of success are mixed.

Following this introduction is a section on background that provides some perspective for the work and discusses some relevant issues. Section 3 contains descriptions of the data acquired for the building of the models developed and treated. Section 4 reports on the model building and testing. The summary in Section 5 includes an outline of the data

structures needed to pursue these issues properly. Compilation of details and other support are in the Appendices.

2. Background

The Navy operates many schools. A main goal of the training system is to place appropriately qualified sailors into the fleet in a timely fashion and in the proper numbers. The planning models no doubt provide for a cushion of reserve in time and personnel, but such planning does not always result in full staffing and the resultant shortfall is certainly always expensive. The extent of the problem is well covered in the references.

There is a basic awkwardness in planning for new recruits to get into boot camp. The recruiting system allows for remarkable flexibility for recruits in terms of entrance times and choice of skill schools. There are many delays charged to AI because of the timing mismatches and to the over-subscription problems, i.e., more students than classroom seats. Of course, there are costs associated to under subscription as well. The awkwardness is exacerbated because remedial action would involve both the recruiting commands and the training commands. Other forms of AI involve transition from one school to the next, and the delays associated with finding a seat when there is a setback, i.e., either an under achieving student being moved to a different section of the same course which had a later convening date or being placed in a prerequisite course for remedial work.

The AT category of NUI also involves liaison with other commands. The main items here are the cutting of orders and budgeting of Permanent Change of Station (PCS) monies. Again, the retrieval of holding time data is difficult.

The most conspicuous cause in the II category of NUI is that of legal holidays. These are easily anticipated, and it seems unlikely that administrative action will be taken to give relief to this source. The number of days lost due to this source should have low variability. Other forms of II occur at random times and may be treated statistically.

One might view the system as an alternating renewal process. A sailor's sojourn in the schools could be marked as "up" when he is in the UI state, and "down" when in the NUI state. The holding times in the down state are likely to have multi-modal distributions. They are influenced by the reason for entering the down state and the accounting rules for the type of down state entered. For example, when a sailor graduates from a class, the graduation date is fixed and capable of being anticipated. Suppose a change of location is required. He enters the AT version of NUI and it seems that the holding time in this state should be deterministic or have a low variance. Further, the follow-on school and its starting time characteristics are known and can be planned upon. At some point, one would expect a transfer to the AI sub-state, but the rules for this change may not be standardized. The main point is that the successful students who are unhindered by random forms of disturbances can flow through a pipeline of schools in a well-planned way (i.e., essentially deterministic). Unfortunately, the data requirements for studying these flows have not been delineated in a structured way. That is, the students are commodities that flow through a network of many paths. The paths must be partitioned into sets, often called pipelines, and every sailor is assigned a particular pipeline. There is some, but a small amount of, lateral transferability

across the pipelines. There is sharing of courses in the early part of the network structure, but afterwards there are a large number of small flows from one school to several specified next schools, and the dispersion dilutes the numbers of sailors. It appears that the retrieval of this type of data must be generated person by person. Unless such distributions have useful stability, standard Renewal Theory models are not likely to be appropriate. Some interesting related network flow models have been introduced, [Lawphongpanich and Brown, 2000].

3. Description of Data

The personnel in charge of the NITRAS database are very cooperative. However, specialized data requests take time and it is not always possible to obtain exactly what we want. We decided to identify about two-dozen prominent courses, by Course Identification Number (CIN), in terms of total dead time and seek longitudinal data for each. The courses having the more complete data are listed. The Course Data Processing Code (CDP) is also marked. (It can identify course location information, whereas the CIN cannot.) We acquired information on them from 1996 through 1999.

CIN	CDP	CIN	CDP	CIN	CDP
A-800-0013	0133	A-623-0125	622N	B-330-0010	3257
A-202-0014	6668	A-730-0010	619D	C-602-2039	625U
A-100-0139	622L	A-661-0010	333K	C-222-2010	619K
A-041-0010	6400	A-661-0103	333L	P-500-0047	253L
A-500-0014	6102	A-431-0069	0519	C-622-2010	619K
A-100-0138	6672	A-651-0119	618J	C-100-2018	642Z
A-202-0014	6668	A-651-0118	617V	C-100-2020	625B
		A-652-0298	6609		

Table 1. Courses with the more complete data.

Initially, the basic categories, AI, II, AT, are marked as to reason (i.e., administrative, legal, medical, and other) with, in the case of AI, on board prior to convening as well. We are also interested in setbacks and attritions. They are less readily anticipated.

It was determined that some important holding-time data can be acquired without sorting through the individual social security numbers. The courses can be accessed from the time point of their convening. The day-by-day events are recorded. It was decided to concentrate on the holding times from the course beginning until academic setback, as; academic attrition, aa; and interrupted instruction, ii (for reasons other than recognized holidays). At these epochs, the cited numbers may enter substates: presumably the aa's go to AT, the as's to AI and the ii's to II. It is not clear how this type of II differs from AI. We do know that an expense is incurred when students leave the UI state. It is useful for the planner to know how many by course and by type (as, aa, ii), and how deeply into the course the student has progressed when this change of state happens. It might be expected that the ii variable is distributed uniformly over time. But the data does not show this. Moreover, it may occur that a particular sailor may experience several II's during a single course. The setbacks and attritions may be related to the portions of the course attempted. If so, it seems that the course administrators have three options: redesign the affected portions of the course, review the admission requirements for the course, or continue to provide for the expense of placing the student into an NUI state.

Accordingly, we proceed to model these processes. They may be useful in determining if these distributions are stable from year to year, how dependent are they upon the particular course, and does the length of a course present an important effect.

4. Modeling

For each course, the number of students leaving the UI state, $\{Y_t\}$ for $t = 1, 2, \cdots, n$ where n is the length of the course in days, is a non-homogeneous Poisson process with mean value function $\{\lambda_t\}$. The modeling process involves finding a description of the $\{\lambda_t\}$ in terms of a few parameters, testing the adequacy of the fit, and assessing the annual stability of the model. Two classes of models were considered: those of the sigmoid learning-curve type, and the more general step-function type.

It was believed that sigmoid models such as the logistic and Gompertz curves, [Hamilton, 1991] would be successful for this purpose, but such did not seem to occur with regularity. We concocted our own model, also of the sigmoid type, and had some success

$$\lambda_t = A \exp\{-a/t - (b \cdot t)^c\}$$

where A, a, b and c are parameters to be fitted. This function stays close to zero in the early part of a course and rises sharply to a single modal value. Then it tapers off with a long right tail. This function captures the idea that there is little in the way of attrition early in a course, and then things change quickly as the early attritions bunch up. The subsequent tapering captures a reduced amount of attrition as the course continues from there. But success with models of this form was limited.

It was decided to work with simple step functions. That is, the sequence of days is partitioned into k intervals and λ_t is constant on each member of the partition. The result is a step function and, if k is not large, it can capture the temporal behavior of the process. These models are general, coarse and can serve to point the way to classes of smoother models.

The fitting process involves the specification of k by the user, and the estimation of the partition break points by maximum likelihood. A special algorithm was developed to accomplish this, and was executed in a Master's thesis, [Li, 2000]. The effect and results of using this algorithm are tabulated in Appendix A. The goodness of fit is judged by a Chi-square statistic with n-k degrees of freedom.

This worked reasonably well. In fact, $k = 5$ led to reasonable fits in many of the cases. But it did not hold up for all. The value $k = 7$ serves for a number of the others.

Turning to the issue of temporal stability, it would be useful if the annual models could be combined into a single choice of partitions for a course. In support of this goal, a distance function was developed in order to measure the separations of the annual functions from the pooled data four-year mean value function. It is described in Appendix B. The use of the pooled mean value function may be tenable, but such use is a judgement call.

5. Summary and Recommendations

The present work makes a beginning on the problem of anticipating the numbers of sailors that enter a NUI state by means of academic attrition, academic setback, and interrupted instruction in terms of how long they have been in the course. Such losses are not uniformly distributed over the length of the course. There is a low level early in a course; the point of rise to more intense leaving activity is illusive; the use of unimodal models to describe this curve may be tenable; the behavior of the curve appears to vary course by course and year by year. A more careful study of these processes would involve the inclusion of the separation of the course segments according to their dates of convening and the enrollment numbers of each. Then sharper modeling can be applied. Of course, more CIN numbers should be included as well.

A planner would need to know the convening dates and the enrollment numbers for the courses in order to use this type of model effectively. The statistical anticipation of losses from these sources could be used to review the administrative aspects of the courses as well as for budgetary planning.

The study of the larger problem, that of down time distributions, requires an investment in defining the processes and organizing the data sources. A first step in the longitudinal analysis of the NUI time in the pipelines would be to specify the data needs. Presumably there are important classes of well-behaved pipes such as the one in the diagram on the next page. This is a schematic in which the network is a tree. The entrance node on the left indicates the beginning of a set of classes (e.g., boot camp, followed by graduation and moving on to another class). The solid lines mark the UI time and the dashed lines indicate the NUI time between courses. In a perfect world, all of the terminal nodes send sailors to the fleet.

Each path in the tree is a pipeline. The tree structure and the implied monotone time scale as one moves from left to right indicates the nature of important dependencies between pipes. Thus, several pipes have common starting points in time. It is assumed that the sequencing of courses is rigid; that is, one must complete a course prior to enrolling in a following course. A set of pipes of such a structure are superimposed on one another, but with staggered starting times. One tree starts at time t_0, the next at $t_1 > t_0$, and so on. The time scales of the two trees need not be identical, but the second and later trees can provide places to put the setbacks.

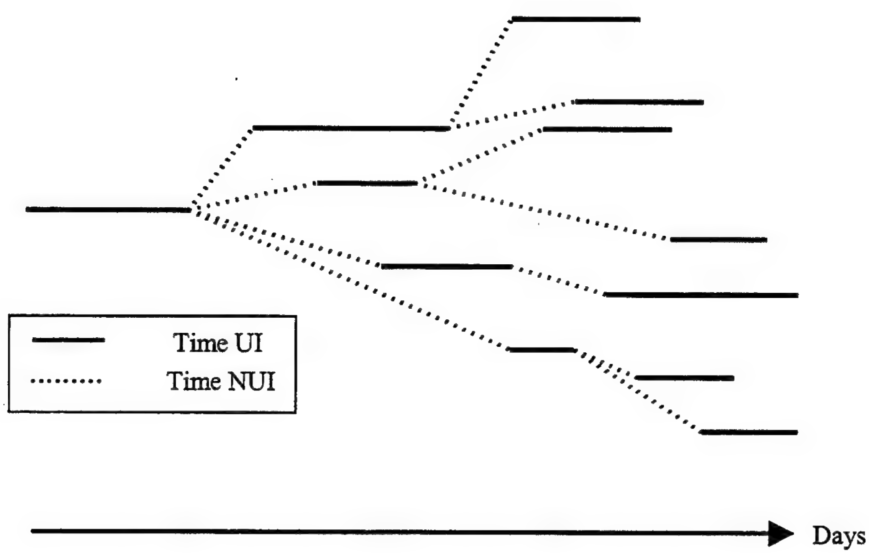


Figure 1. A tree representation of course pipelines.

The analyst needs a set of trees, the course numbers of all courses in a pipe and their capacities, the convening dates, and the enrollment and graduation numbers for each course. When a change of location is involved between courses they should be so marked. With data of this type one can do the following: identify the efficient and inefficient pipes; compute losses at the end of each course and the terminal nodes; determine the holding time distributions; and set priorities for the next set of actions and studies.

It is recommended that:

- A representative set of pipelines be identified, which are well behaved in that they process substantial numbers of sailors and reflect important classes of end-product skills.

- Databases and query systems be generated so that a researcher has convenient access to the information outlined above: convening and graduation dates, enrollment and graduation numbers, course capacities and locations, dates of course attritions in various categories and policies for managing those who do not complete courses as scheduled.

References

Belcher, S. W., Reinert, V. C. and Hiatt, C. M., (1999). "Analysis of Student Not-Under-Instruction Time in Initial Skills Training: Trends, Causes, and Proposed Fixes," Center for Naval Analysis, Alexandria, VA.

Lawphongpanich, S. and Brown, R., (2000). "Determining the Onboard Capacity for Initial Skill Training," Operations Research Department, Naval Postgraduate School, NPS-OR-01-003-TR

MacIlvaine, M. E., (1998). "Street-To-Fleet Dataset, Center for Naval Analysis, Alexandria, VA.

Hamilton, L. C., (1991). *Regression with Graphics*, Duxbury.

Li, J., (2000). "A Poisson Regression Analysis of the Academic Setback in Naval Training Dead time," Master's Thesis, Systems Management Department, Naval Postgraduate School.

Sladyk, Rudy, (1999). "Recruit Training Pipelines: RTC to Fleet," Briefing slides - 3 March, N813G

Quester, A.O., MacIlvaine, M. E., Barfield, L. C., Parker, L. J., and Reese, D. L. (1998). "Final Report for CNA Study on Answering Decision-Makers Questions: Organizing Information for Policy Analysis," Center for Naval Analysis , Alexandria, VA.

Rhoades, K. M., (1999). "Dead Time in Naval Training," Master's Thesis, University of Maryland.

Appendix A: Model Fitting Summaries

The tables serve to illustrate the results of model fitting and the amount of variability. This is done for the temporally combined data and for the individual years in selected cases. The use of five intervals in a partition is often acceptable, but there are courses and lost time types for which five is inadequate from a statistical point of view. The model fitting is by maximum likelihood; the estimates are the partition break points and the mean value rates. The number of partitions, k, is user supplied.

The individual years bear but small resemblance to each other and to the combined years. Three types—academic attrition, academic setback, and interrupted instruction—do not appear to behave in common patterns.

Legend: b: the partition break points, days since inception.
p: the length of the interval, days.
λ: the estimated rate for the interval.
Y: the total number of events in the interval.
The length of the course is the last entry in the b column.

Academic Attrition

622L Combined				1996				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
20	20	0.00	0	20	20	0.00	0	50	50	0.00	0	41	41	0.00	0	34	34	0.06	2
34	14	0.36	5	22	2	1.50	3	52	2	1.00	2	72	31	0.39	4	44	10	1.90	19
90	56	2.62	147	34	12	0.00	0	93	41	0.12	5	135	62	0.75	47	48	4	0.25	1
128	38	4.00	152	112	78	0.74	58	106	13	1.00	13	141	6	0.00	0	59	11	2.36	26
141	13	2.38	31	141	29	1.59	46	141	35	0.49	17					141	82	0.98	80

619K Combined				1996				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
10	10	0.30	3	20	20	0.25	5	34	34	0.26	9	10	10	0.00	0	12	12	0.00	0
36	26	3.35	87	46	26	1.15	30	42	8	2.25	18	28	18	1.50	27	53	41	1.15	47
45	9	6.56	59	67	21	0.10	2	85	43	0.63	27	34	6	0.17	1	84	31	0.39	12
86	41	1.88	77	71	4	1.25	5	96	11	0.00	0	51	17	1.41	24	91	7	0.00	0
110	24	1.04	25	110	39	0.18	7	110	14	0.57	8	110	59	0.39	23	110	19	0.32	6

618J Combined				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
26	26	0.00	0	26	26	0.00	0	31	31	0.00	0	35	35	0.00	0
51	25	0.76	19	55	29	0.14	4	55	24	0.54	13	56	21	0.10	2
55	4	0.00	0	57	2	2.00	4	60	5	1.40	7	67	11	0.36	4
57	2	5.00	10	62	5	0.00	0	66	6	0.00	0	70	3	1.67	5
81	24	1.25	30	81	19	0.37	7	81	15	0.67	10	81	11	0.27	3

6668 Combined				6609 Combined			
b	p	λ	Y.	b	p	λ	Y.
24	24	0.04	1	21	21	0.00	0
52	28	0.61	17	34	13	0.46	6
54	2	3.00	6	50	16	0.12	2
56	2	0.00	0	51	1	2.00	2
96	40	0.95	38	56	5	0.00	0

Academic Setbacks

622L Combined				1996				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y
13	13	2.00	26	22	22	2.00	44	13	13	0.08	1	41	41	0.00	0	14	14	0.14	2
30	17	35.71	607	29	7	22.71	159	29	16	13.81	221	72	31	0.39	12	80	66	6.61	436
41	11	7.18	79	57	28	2.18	61	40	11	2.36	26	73	1	4.00	4	98	18	3.33	60
106	65	28.11	1827	104	47	8.66	407	107	67	10.11	677	135	62	0.76	47	99	1	28.00	28
141	35	10.00	350	141	37	3.00	111	141	34	3.12	106	141	6	0.00	0	141	42	1.40	59

619K Combined				1996				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
1	1	48.00	48	1	1	20.00	20	1	1	28.00	28	10	10	0.10	1	8	8	0.00	0
8	7	0.43	3	6	5	0.00	0	13	2	0.33	4	16	6	6.83	41	42	34	2.85	97
38	30	10.60	318	17	11	3.18	35	36	3	3.00	69	38	22	2.50	55	68	26	0.65	17
79	41	2.56	105	35	18	1.44	26	59	3	0.00	0	91	53	0.58	31	70	2	3.50	7
110	31	0.68	21	110	75	0.29	22	110	51	0.55	28	110	19	0.00	0	110	40	0.35	14

618J Combined				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
14	14	0.57	8	14	14	0.14	2	20	20	0.55	11	20	20	0.40	8
27	13	6.92	90	25	11	1.64	18	27	7	3.29	23	27	7	3.14	22
29	2	39.00	78	29	4	8.50	34	29	2	17.00	34	29	2	12.00	24
54	25	10.28	257	62	33	3.15	104	52	23	3.48	80	54	25	3.44	86
81	27	3.26	88	81	19	0.74	14	81	29	1.17	34	81	27	1.00	27

Interrupted Instruction

622L Combined				1996				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
7	7	14.71	103	7	7	5.14	36	13	13	6.08	79	27	27	4.56	123	85	85	6.47	550
8	1	111.00	111	8	1	88.00	88	47	34	11.71	398	77	50	8.54	427	86	1	103.00	103
85	77	36.212	788	69	61	8.10	494	90	43	15.65	673	78	1	31.00	31	92	6	7.50	45
86	1	156.00	156	77	8	3.62	29	92	2	35.00	70	126	48	8.69	417	93	1	27.00	127
141	55	54.78	3013	141	64	11.97	766	141	49	14.45	708	141	15	7.53	113	141	48	18.62	894

619K Combined				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
79	79	5.04	398	4	4	6.50	26	14	14	0.71	10	79	79	1.66	131
80	1	30.00	30	17	13	2.08	27	15	1	8.00	8	80	1	30.00	30
86	6	1.50	9	24	7	0.43	3	60	45	1.60	72	86	6	0.00	0
87	1	23.00	23	38	14	3.21	45	61	1	10.00	10	87	1	20.00	20
110	23	4.35	100	110	72	0.82	59	110	49	1.12	55	110	23	2.78	64

618J Combined				1996				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
17	17	6.53	111	10	10	1.1	11	35	35	3.97	139	18	18	0.89	16	4	4	0.50	2
35	18	11.11	200	22	12	0.0	0	36	1	26.00	26	67	49	2.33	114	56	52	2.65	138
36	1	34.00	34	64	42	1.6	67	54	18	7.67	138	69	2	0.00	0	57	1	8.00	8
57	21	14.90	313	79	15	0.4	6	55	1	26.00	26	74	5	3.40	17	70	13	1.08	14
81	24	10.62	255	81	2	4.0	8	81	26	5.31	138	81	7	0.86	6	81	11	3.55	39

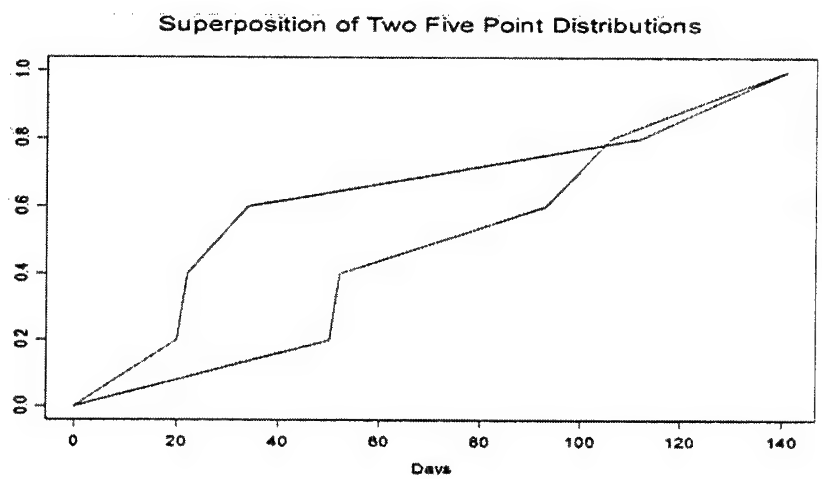
6668 Combined				1997				1998				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
6	6	1.83	11	11	11	1.09	12	13	13	0.69	9	11	11	1.55	17
35	29	7.34	213	12	1	12.00	12	42	29	2.66	77	12	1	9.00	9
40	5	15.40	77	82	70	2.17	152	46	4	6.75	27	20	8	0.38	3
41	1	0.00	0	91	9	5.44	49	48	2	0.00	0	88	68	4.84	329
96	55	12.33	678	96	5	1.80	9	96	48	4.85	233	96	8	5.12	41

6609 Combined				1996				1997				1999			
b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.	b	p	λ	Y.
1	1	0.00	0	7	7	2.43	17	3	3	0.00	0	4	4	0.00	0
7	6	9.83	59	9	2	20.00	40	8	5	8.40	42	40	36	2.06	74
10	3	20.67	62	28	19	3.32	63	10	2	0.00	0	43	3	0.33	1
12	2	33.00	66	29	1	20.00	20	12	2	25.00	50	44	1	8.00	8
56	44	7.80	343	56	27	1.70	46	56	44	3.59	158	56	12	0.92	11
56	56	5.00	530												

Appendix B: Measuring the Distance Between Two Models

The class of models is the family of simple step functions. These models are precursors to smooth curve models that describe the mean value function of the non-homogeneous Poisson processes that describe the attrition/setback/interruption events that occur in the time period of a course. The distance function chosen is one that is compatible with this more encompassing class of models. The step functions are treated as densities, and the distance between two such functions is the integral of the magnitudes of the differences separating their cumulative distribution functions.

The graph below will illustrate the point. The two models are described by their partition points (i.e., column b of the tables in Appendix A). When k=5, this is viewed as a distribution over five points. One forms the cumulative distribution values at the epochs of change and connects the points with straight-line segments. The graph shows this for the 1996 and 1997 partition distributions of course 622L for academic attritions. The course length is 141 days and each model has a partition of five break points. The distance between the two is the magnitude of the areas separating them, measured as a percentage of the area of the containing rectangle. The separation of the two curves shows great year-to-year variability. The code for computing this distance is in Appendix C.



The following sets of distance tables provide an image of the distances between the models for various years with a single course and event type. The column marked "all" refers to the model that combines the data for all of the years. The distances in the "all" column are not generally smaller than the inter-year distances.

Academic Attrition.

	all	96	97	98
	0			
96	11.9	0		
97	10.0	17.0	0	
98	9.4	18.9	7.1	0
99	18.0	13.0	16.5	19.3

Academic Setback

	all	96	97	98
	0			
96	3.5	0		
97	0.4	3.8	0	
98	18.6	15.5	18.7	0
99	15.4	13.3	15.7	10.6

Interrupted Instruction

	all	96	97	98
	0			
96	3.5	0		
97	7.9	11.5	0	
98	17.5	20.9	10.3	0
99	24.1	27.7	16.2	14.8

CDP 619K

Academic Attrition.

	all	96	97	98	
	0				
96	8.7	0			
97	14.5	9.9	0		
98	9.8	14.7	24.4	0	
99	11.5	8.8	5.6	21.3	0

Academic Setback

	all	96	97	98	
	0				
96	12.2	0			
97	4.6	9.	0		
98	5.3	17.5	8.4	0	
99	13.3	23.5	14.4	11.8	0

Interrupted Instruction

	all	97	98	
	0			
97	45.3	0		
98	33.1	12.	0	
99	0.0	45.3	33.1	0

CDP 618J

Academic Attrition.

	all	97	98	
	0			
97	2.7	0		
98	5.7	3.0	0	
99	9.6	6.9	4	0

Academic Setback

	all	97	98	
	0			
97	2.5	0		
98	2.0	4.4	0	
99	1.5	4.0	0.5	0

Interrupted Instruction

	all	96	97	98	
	0				
96	15.1	0			
97	9.2	16.6	0		
98	20.5	14.9	17.	0	
99	14.8	11.1	14.0	10.1	0

Appendix C: S-Plus Code for Computing Distances Between cdf's

The first function, seg.comp(), computes the area of a polygon marked by crossover points of the two cdf's. It is signed by the order of the input. The second function, area.comp(), collects all of the signed area segments of the two in the first column of the output. The other columns contain information useful in more extensive applications. The third function, dist.mat(), develops the lower triangular distance matrix for a collection of models, each column of the input matrix is the set of partition points for a model. There is also an auxiliary program, sol.pt().

```
seg.comp
function(x, w, u0, y0, n0)
{
# fname is seg.comp
# Computes the areas under the polygonal curves, between
# two knots, and returns their difference (signed). A flag
# is set = 1 if the x cdf is above the w cdf, and set = 2
# otherwise. The x and w vectors are mono increasing; n0 is
# the number of points in the original full sets. The initial points
# (u0, y0) mark the beginning of the segment; the cross-over
# point (u1,y1) is the segment end and is computed internally. A
# special adjustment is made if there are no crossover points.
        ss <- sort(c(x, w))
        n <- length(x)
        flag <- 1
        if(w[1] == ss[1])
                flag <- 2
        j <- 1:n
        dx <- diff(x)
        dw <- diff(w)
        if(flag == 1 & sum(w[j] >= x[j]) == n) {
                area1 <- 0.5 * (y0 + 1) * (x[1] - u0) + dx %*% (j[ - n]                + 0.5) + n
* (w[n] - x[n])
                area2 <- 0.5 * (y0 + 1) * (w[1] - u0) + dw %*% (j[ - n]                + 0.5)
                u1 <- x[n]
                y1 <- 0
                f <- n}
```

```
        else if(flag == 2 & sum(w[j] <= x[j]) == n) {
                area1 <- 0.5 * (y0 + 1) * (x[1] - u0) + dx %*% (j[ - n]          + 0.5)
                area2 <- 0.5 * (y0 + 1) * (w[1] - u0) + dw %*% (j[ - n]              + 0.5) +
n * (x[n] - w[n])
                u1 <- w[n]
                y1 <- 0
                f <- n}
        else {
                ind <- j[x[j] >= w[j]]
                if(flag == 2)
                        ind <- j[w[j] >= x[j]]
                f <- ind[1]
                if(f == 1) {
                        area1 <- 0
                        area2 <- 0
                        u1 <- x[1]
                        y1 <- 0}
                if(f > 1) {
# make the end corrections.
                        P1 <- c(x[(f - 1):f])
                        P2 <- c(w[(f - 1):f])
                        tout <- sol.pt(P1, P2)
                        u1 <- tout[1]
                        y1 <- tout[2] + f - 1
                area1 <- ((x[1] - u0) * (1 + y0))/2 + ((u1 - x[f - 1])           * (y1 + f -
1))/2
                area2 <- ((w[1] - u0) * (1 + y0))/2 + ((u1 - w[f - 1])           * (y1 + f -
1))/2}
                adj1 <- adj2 <- 0       # initialize the adjustments in the center
                if(f >= 3) {
#       adj1 <- (x[f - 1] * (2 * f - 3))/2 - (3 * x[1])/2
#     adj2 <- (w[f - 1] * (2 * f - 3))/2 - (3 * w[1])/2
#  }
#  if(f > 3) {
#       adj1 <- adj1 - sum(x[2:(f - 2)])
#       adj2 <- adj2 - sum(w[2:(f - 2)])
                        j <- 2:(f - 1)
                        dx <- diff(x[1:(f - 1)])
                        dw <- diff(w[1:(f - 1)])
                        adj1 <- dx %*% (j - 0.5)
                        adj2 <- dw %*% (j - 0.5)}
```

```
                y1 <- y1 - f + 1
                area1 <- area1 + adj1
                area2 <- area2 + adj2}
        net <- (area1 - area2)/n0
        out <- c(net, flag, f, u1, y1)
        names(out) <- c("net", "flag", "f", "u1", "y1")
        out}

area.comp
function(x, w, u0 = 0, y0 = 0)
{# fname is area.comp
# Computes the signed net areas separating the empirical
# cdf's of the ordered sets x and w.  These cdf's are polygonal
# curves which are connected with straight line segments.  The
# two data sets are of the same length.
        out <- NULL
        n0 <- length(x)
        jj <- 1
        repeat {
                out <- rbind(out, seg.comp(x, w, u0, y0, n0))
                assign("out", out, frame = 0)
                f <- out[jj, 3]
                u0 <- out[jj, 4]
                y0 <- out[jj, 5]
                x <- x[-1:(1 - f)]
                w <- w[-1:(1 - f)]
                n <- length(w)
                if(n == 1)
                        break
                jj <- jj + 1}
        out}

dist.mat
function(mat)
{
# fname is dist.mat
# Computes the distance between models of a course
# for the several years.  Input is matrix whose columns are the fitted
# models.  Output is lower triangular and in the percent of the area
# of the rectangle.
        n <- ncol(mat)
        n0 <- nrow(mat)
        dd <- matrix(0, n, n)
        for(i in 1:(n - 1)) {
```

```
		j <- i + 1
		repeat {
			tmp <- area.comp(mat[, i], mat[, j])
			dd[j, i] <- sum(abs(tmp[, 1]) * 100)/mat[n0, 1]
			if(j == n)
				break
			j <- j + 1}}
	dd <- round(dd, 1)
	dd}

sol.pt
function(P1, P2){
# fname is sol.pt
# finds the cross over solution point
# for two cdf's that have the same number
# of pts in the horiz & equi spaced in the vert.
	x1 <- P1[1]
	x2 <- P1[2]
	w1 <- P2[1]
	w2 <- P2[2]
	delx <- x2 - x1
	delw <- w2 - w1
	x <- (x1 * w2 - w1 * x2)/(delw - delx)
	y <- (x - x1)/delx
	out <- c(x, y)
	out}
```

INITIAL DISTRIBUTION LIST

1. N8, Office of the Chief of Naval Operations 2
   2000 Navy Pentagon
   Washington, DC 20350-2000

2. Defense Technical Information Center 2
   8725 John J. Kingman Road, STE 0944
   Ft. Belvoir, VA 22060-6218

3. Research Office (Code 09) 1
   Naval Postgraduate School
   Monterey, CA 93943-5000

4. Dudley Knox Library (Code 013) 2
   Naval Postgraduate School
   Monterey, CA 93943-5002

5. Prof. Robert R. Read (Code OR/Re) 5
   Department of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000

6. Richard Mastowski (Editorial Assistant) 2
   Department of Operations Research
   Naval Postgraduate School
   Monterey, CA 93943-5000